Cross-Cutting Science: Paving the Way to Discovery

Advances in medicine are largely dependent on the accumulation of new knowledge about biologic processes, especially at the smallest levels of an organism—its genes, the proteins they control, and the workings of cells. The ultimate application of such very basic research is not always obvious, either to the public or to the investigator. However, major strides in fighting disease can be traced back to laboratory studies whose immediate relevance to health could not have been fully known or appreciated at the time they were conducted. Described here are some recent studies of fundamental processes, as well as the technologies that make such studies possible. The critical insights gained through this type of research can be expected to propel disease-oriented research, not only within the NIDDK mission, but in many other fields. Investment in such cross-cutting scientific research today will have future applications that we cannot now describe with certainty, but which we know will surely be realized.

Genetic Imprinting

The genome is the set of genetic instructions within the strands of DNA, the genetic blueprint in each cell's nucleus. It is the genome that determines not only if a human being is created but how that person will develop. How does the genome do this? What determines which of the genes along this strand will be activated or "expressed" in the cell in order to dictate that cell's functional capabilities? Usually, each of the genes a person inherits comes in two copies, one from the father and one from the mother. Traditional genetics predicts that both copies of the gene will function the same way. A striking departure from this prevailing principle of genetics is a process called "imprinting," in which the genes from only one parent's chromosome are expressed. One of the two copies of a gene is inactivated or "switched off," while the other is activated or "switched on." Researchers are seeking to comprehend this "switching" mechanism, through which a gene's coded information is activated and expressed—first, at the cellular level, but ultimately, in human traits and characteristics—in health and disease. They have suggested that imprinting protects organisms that would otherwise be harmed if both copies of the gene were actively producing proteins in a "double dose." In fact, abnormalities in the imprinting process are often associated with human diseases, including disorders affecting cell growth, cell development, and cell function.

For example, in the production of some growth factors, over-expression of certain genes in humans is associated with fetal overgrowth and, in some cases, the development of tumors. This occurs with the gene for insulin-like growth factor 2. When imprinting happens, the expressed gene may come from either the mother or the father. Whether the maternal or paternal gene is beneficially "silenced" in imprinting depends on the presence of markers, called methyl groups, that chemically modify the DNA near the gene. In the case of insulin-like growth factor 2, the copy inherited from the father is functional and has methylated DNA nearby, whereas the copy from the mother is unmethylated and inactive. However, if the maternal copy of the gene is inappropriately methylated, there is over-expression of the protein, which is associated with a disorder that predisposes children to certain malignancies. A cutting-edge research area includes studies to determine what factors play a role in a wide range of diseases such as this, in which gene expression is abnormal.

This is a computer-generated ribbon diagram of the three-dimensional structure of a protein-DNA complex. The protein shown in this image is an enzyme responsible for detecting errors in the DNA sequence. The enzyme, called MutS, has several different domains, shown in dark blue, light blue, yellow, and orange. A strand of DNA is shown in red and is surrounded by the domains of the MutS enzyme. Scientists in the field of structural biology study the shape of proteins to look for clues about how they perform their critical functions. Knowing the shape of molecules can also help scientists design drugs. Photo: Dr. Wei Yang, Laboratory of Molecular Biology, NIDDK.

For insulin-like growth factor 2, scientists have discovered insulators of gene expression that may be critically important to the beneficial "silencing" of the gene. Investigators have reported that a certain protein, called CTCF, acts to prevent the gene from making the growth factor, which in excessive quantities is linked to malignancy. The research team had been studying this protein because it binds to a "boundary" element between the loosely configured DNA, where genes are active, and the more tightly packed DNA, where genes are inactive. Previous work had shown that the CTCF protein binds to specific sites on DNA and insulates, or blocks, gene-activating elements called "enhancers" from turning on the genes they help activate. Based on these data, the group hypothesized that the CTCF protein played a similar role in ensuring that the maternal copy of the insulin-like growth factor 2 gene remained switched off and that harmful quantities of the growth factor were not produced. The researchers then conducted a series of experiments to test their hypothesis. Further studies showed that DNA near the insulin-like growth factor 2 genes of mice and humans did indeed contain binding sites for this insulating protein. Using cells in culture, researchers next showed that these binding sites can only block the action of an "enhancer" on the gene when the binding sites lie between the enhancer and gene, a signature of insulator activity. In the case of the insulin-like growth factor 2 gene, the protein binding sites are found in the same region of DNA that is usually methylated on the paternal copy of the gene and unmethylated on the maternal copy. In the final experiment, the researchers showed that adding methyl groups to the DNA binding sites for this protein prevents the protein from binding. This finding strongly suggests that methylation abolishes the insulating capability of the protein on the paternal copy of the gene. As a result, the enhancer can activate paternal gene expression and appropriate quantities of insulin-like growth factor 2 can be produced.

Scientists have long recognized the importance of methylation in imprinting, but many other mechanisms governing imprinting remain unknown. A leading investigator in the field had hypothesized earlier that the insulin-like growth factor 2 region contained an insulator that prevented the maternally inherited gene from expressing growth factor and that methylation blocked this insulator. The research studies of the protein CTCF confirmed this hypothesis. More importantly, the studies demonstrated that the control mechanism for insulating against the activity of this imprinted gene is a novel "boundary" element, which the investigators are calling the "imprinting control region." Also, the fact that protein binding, and therefore insulator activity, can be turned "on" and "off" by changing the methylation state of DNA suggests that there may be many other places in the genome at which an insulator suppresses expression of a gene. Thus, these research findings could have broad application to other diseases of abnormal gene expression.

Because imprinting is a departure from traditional principles of genetics, it offers a model for understanding how modifications in DNA structure can regulate gene expression. Clearly, disruptions in the genetic programs controlling the regulation of genes and proteins, as well as mutational alterations in the genes, can have many disease-causing effects on living organisms. For example, imprinting plays an important role in certain endocrine or hormonal disorders. When proteins responsible for conveying a signal from a hormone receptor to machinery inside the cell are nonfunctional, the signal is lost and the patient develops diseases of hormone regulation. The hereditary pattern of some hormone regulatory diseases suggests that they are caused by errors in imprinting.

A complete understanding of imprinting is also essential to deciphering the information contained in the human genome. Insights will emerge from studies of the structure and function of insulators and related boundary elements to determine if methylation is a key process. A practical implication of this new knowledge of how insulators regulate gene expression will be its application to the design of "vectors" for use in delivering healthy genes to repair the damage done by abnormal ones (see the following research advance). New approaches are needed that overcome problems with low gene expression associated with conventional vector design. The NIDDK continues to support vigorous research on gene expression, gene transfer, and gene vector development, both extramurally and intramurally, in order to enhance and apply emerging knowledge of the human genome.

Bell AC and Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *IGF2* gene. *Nature* 405(6785):482-485, 2000.

Bell AC, West AG, and Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98(3):387-396, 1999.

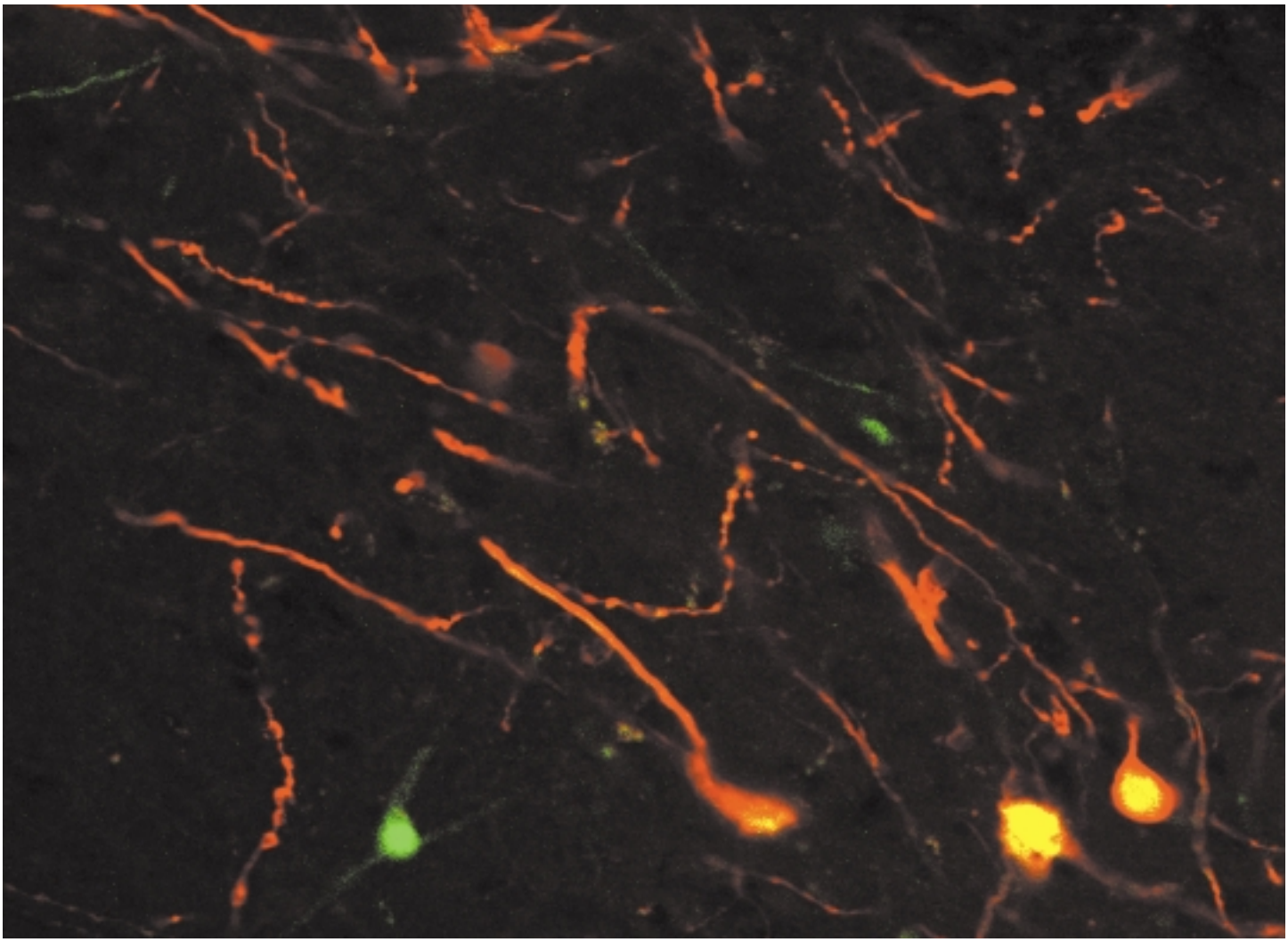
Recillas-Targa F, Bell AC, and Felsenfeld G. Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proc Nat Acad Sci USA* 96(25):14354-14359, 1999.

Saitoh N, Bell AC, Recillas-Targa F, West AG, Simpson M, Pikaart M, and Felsenfeld G. Structural and functional conservation at the boundaries of the chicken beta-globin domain. *EMBO J* 19(10):2315-2322, 2000.

Fighting Disease by Delivering Healthy Genes

The day-to-day activities of a living cell largely depend on the proper function of many cellular proteins. Proteins are formed based on the information contained within a person's genes. Thus, when the genetic information contains an error, or mutation, the protein may be formed incorrectly. If a protein is malformed, it can't do its job well and the activities of the cell are disrupted. For example, in sickle cell disease, malformed red blood cell proteins cause the cell to be sickle-shaped, which impairs its ability to transport oxygen.

In order to treat patients whose disease results from a genetic error, scientists are investigating the therapeutic potential of a technique called "gene transfer." In this approach, the patient is injected with a "healthy" copy of the genetic information, to compensate for or replace the abnormal copy in the patient's own cells. It is hoped that the "healthy" copy will make normal proteins, and thus eliminate or mitigate the harmful consequences of the disease. Many genes have already been identified and isolated—such as the genes responsible for hereditary



Scientists are studying the use of gene transfer techniques to deliver healthy genes to many types of body tissues. One delivery method uses viruses to carry the gene to the target tissue. Here, neurons in the rat brain were infected with two adeno-associated viruses (AAVs) engineered to contain a gene for either a red or green fluorescent protein. Cells fluorescing yellow contain both viruses. Photo: Dr. Krys Bankiewicz, University of California San Francisco, and Dr. Michael Schmidt, National Heart, Lung and Blood Institute, National Institutes of Health.

pancreatitis and cystic fibrosis—but the technology necessary to use healthy copies of them for therapeutic gene transfer is not fully refined.

An obstacle in gene transfer research is the limitation of the carrier or "vector" used to transport the genetic information into a patient's cells. Most vectors are derived from viruses whose disease-causing genes are replaced with one or more therapeutic genes to render the virus into a harmless but efficient transport vehicle. Viral vectors are useful for gene therapy because intact portions of the viral genes enable them to make numerous copies of themselves. Problems arise, however, when the human body recognizes these viral "carrying-cases" as "foreign." Then, the immune system eventually destroys the viral vectors and the therapeutic genes contained within them. As a result, scientists have been searching for alternative vectors that do not initiate an immune response. One such vector is the adeno-associated virus, or AAV. This vector has the demonstrated ability to deliver genes to many different kinds of body tissues and to maintain their expression for a prolonged time. Unfortunately, this vector has its own drawback. The virus itself is small, therefore inserted therapeutic genes must also be small. Until now, this problem has limited the use of the AAV vector to studies of disorders whose mutated gene is small in size. However, three independent groups of researchers recently discovered that they could break a therapeutic gene into two parts, put each part into an AAV, and deliver the parts in tandem. The two halves of the gene join together prior to protein production, and thus, the end product is the full-sized protein. Using this method, scientists were able to successfully treat a mouse model of hemophilia.

In the search for better vectors, another group of NIDDK-supported scientists has developed a non-viral vector for delivering therapeutic genes. This method uses what the scientists call "naked" DNA—called "naked" because it is not contained within a virus that would be subject to an immune-system attack. Included along with the naked DNA is an element known as a transposon—a naturally occurring part of a chromosome that is capable of moving from one location to another. To change locations, transposons move to a new position, cut the chromosome, and paste their DNA into that position. Importantly, transposons can be manipulated so that they insert therapeutic genes into the chromosome along with their own DNA. This feature makes them exceptionally attractive "carrying-cases" for gene transfer. Moreover, because the therapeutic gene is inserted into the chromosome, it can potentially be expressed for a long time. Using naked DNA coding for a blood-clotting factor combined with a transposon, scientists have been able to induce blood clotting in a mouse model of hemophilia.

One day, many complications of diabetes and other complex genetic diseases may also be amenable to treatment with gene transfer—as the causative genes are identified and ways are found to replace or repair them. However, limitations in vectors and related techniques are impeding the pace of progress in the therapeutic use of gene discovery. To counter these problems, the NIDDK is promoting research on the development of new vectors that could be used to treat any disease caused by a genetic mutation. The Institute's emphasis on vector development complements an extensive portfolio of ongoing basic research in molecular genetics, and research centers programs in areas including molecular hematology, cystic fibrosis, and polycystic kidney disease.

An obstacle in gene transfer research is the limitation of the carrier or "vector" used to transport the genetic information into a patient's cells.

Duan D, Yue Y, Yan Z, and Engelhardt JF. A new dual-vector approach to enhance recombinant adeno-associated virus-mediated gene expression through intermolecular *cis* activation. *Nat Med* 6(5):595-598, 2000.

Nakai H, Storm TA, and Kay MA. Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. *Nat Biotechnol* 18(5):527-532, 2000.

Samulski RJ. Expanding the AAV package. *Nat Biotechnol* 18(5):497-498, 2000.

Sun L, Li J, and Xiao X. Overcoming adeno-associated virus vector size limitation through viral DNA heterodimerization. *Nat Med* 6(5):599-602, 2000.

Yan Z, Zhang Y, Duan D, and Engelhardt JF. Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proc Natl Acad Sci USA* 97(12):6716-6721, 2000.

Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, and Kay MA. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat Genet* 25(1):35-41, 2000.

Cell-Based Approaches to Therapy

Studies of cell and organ development are generating a wealth of knowledge from which new approaches to treating and preventing disease can be expected to emerge. One research focus is the body's stem cells, which are found at the very earliest stages of development. Stem cells are undifferentiated progenitor cells, which ultimately specialize into blood, heart, kidney, liver, eye and other cells of the body. However, even as adults, human beings retain some stem cells, particularly in bone marrow and the gut lining, and possibly also in other tissues. These mature stem cells, often called "adult" stem cells, hold therapeutic potential because of the possibility that they could be "coaxed" to differentiate into a broad range of organ-specific cells that could then be used to treat or cure disease.

Several recent studies have discovered new knowledge about stem cells of the bone marrow and fetal liver, termed "hematopoietic" stem cells. In a study in mice, researchers used specific markers to separate all of the hematopoietic and lymphoid progenitor cells from bone marrow cells, and then studied the remaining sub-populations of cells. In this way, they found a common myeloid progenitor cell in the mouse—a discovery that will help to advance stem cell research. In related work, researchers reported using a combination of analytic techniques to identify novel stem cell genes in fetal mouse liver tissue. From extracted cells, they identified genetic factors involved in the direction of protein synthesis. They also identified cell-surface molecules, secreted proteins and signaling molecules. Many of these were previously undescribed or were found in stem cells for the first time. These types of studies set the stage for future investigations of the transformational events leading to the multitude of blood disorders, and more broadly, for investigations of genes involved in the commitment of cells to specific lineages. Moreover, the discovery of stem cell proteins will facilitate studies of networks of protein interactions, a field known as proteomics. This knowledge will likely be critical for future research that can produce specialized cells for treating disease.

Working with the idea that hematopoietic stem cells are able to form liver tissue, researchers have gained supportive evidence from a mouse model with a potentially lethal genetic defect in the liver. These animals are deficient in the enzyme FAH and die shortly after birth unless treated. To test their theory, the researchers destroyed the bone marrow in the FAH-deficient mice through irradiation, then gave the mice a bone marrow transplant. The transplanted cells had the ability to make FAH and they also contained a marker enzyme, so that they and their progeny could be identified. Remarkably, the experimental animals no longer needed treatment and remained healthy. Autopsy of the animals confirmed that new liver cells had indeed been formed from the transplanted hematopoietic stem cells. The researchers then proceeded to repeat the experiment with cells separated from the total bone marrow population based on their surface markers. These experiments enabled the scientists to identify the specific cell population in bone marrow that was responsible for replacing the hematopoietic function and repairing the defective liver function. This research is exciting because it suggests ways to correct abnormal organ function due to genetic defects or disease.

Studies in humans have also confirmed that bone marrow cells can migrate to the liver and differentiate into mature functional liver cells. Researchers analyzed tissue samples obtained from women who had received a bone marrow transplant from a male donor. They identified liver cells containing a Y chromosome that could only have originated from the transplant because women have no Y chromosomes. Similarly, researchers identified liver cells containing a Y chromosome in a male patient who had received a liver transplant from a woman. The woman's liver contained only X chromosomes; therefore, the man's own cells must have migrated to the new liver received from the woman. The demonstration that mature liver cells can be derived from circulating cells, most likely of bone marrow origin, has major implications for treatment of liver failure and genetic liver diseases via liver-cell transplantation and gene transfer approaches. It also has widespread implications for the use of adult stem cells in treating other diseases.

Progress in type 1 diabetes may also be attained through stem-cell research. Scientists are urgently hunting for adult stem cells in the pancreas—cells that might be used for restoring insulin-producing capacity to patients with this disease. In laboratory experiments, investigators have generated insulin-producing islets, from cells isolated from the pancreatic ducts of non-obese diabetic mice, before they developed diabetes. When they implanted the islets into the kidney of non-obese diabetic

mice, the researchers were able to wean the mice off insulin injections. The mice also exhibited a reduction in blood glucose levels, which approached normal levels. This research is very promising; however, future studies will be necessary to verify that the reduction in glucose levels seen in these experiments was the result of the islets implanted into the kidney. On a parallel track, other NIDDK-supported investigators have shown that islet-like clusters of pancreatic endocrine cells can be cultured from pancreatic tissue that is usually discarded after isolation of islets for research studies. The islet-like clusters were capable of producing insulin in response to glucose in culture.

Collectively, these recent studies fuel hope for the possible use of controlled differentiation of stem cells to obtain specialized cells for treating or reversing many diseases. The NIDDK is pursuing this line of research through a new initiative to encourage investigators to explore the capability of stem cells to differentiate–a characteristic known as "plasticity." The Institute is also establishing a Beta Cell Biology Consortium to investigate every aspect of insulin-producing cells for clues to the treatment and prevention of type 1 and type 2 diabetes. Other initiatives are likely to emerge from a newly formed NIDDK Strategic Planning Group, which will make recommendations regarding research leads in stem cell biology and general developmental biology.

Bonner-Weir S, Taneja M, Weir GC, Tatarkiewicz K, Song KH, Sharma A, and O'Neil JJ. In vitro cultivation of human islets from expanded ductal tissue. *Proc Natl Acad Sci USA* 97(14):7999-8004, 2000.

Lagasse E, Connors H, Al-Dhalimy M, Reitsma M, Dohse M, Osborne L, Wang X, Finegold M, Weissman IL, and Grompe M. Purified hematopoietic stem cells can differentiate into hepatocytes *in vivo*. *Nat Med* 6(11):1229-1234, 2000.

Orkin SH. Stem cell alchemy. *Nat Med* 6(11):1212-1213, 2000.

Ramiya VK, Maraist M, Arfors KE, Schatz DA, Peck AB, and Cornelius JG. Reversal of insulin-dependent diabetes using islets generated in vitro from pancreatic stem cells. *Nat Med* 6(3):278-282, 2000.

Theise ND, Nimmakayalu M, Gardner R, Illei PB, Morgan G, Teperman L, Henegariu O, and Krause DS. Liver from bone marrow in humans. *Hepatology* 32(1):11-16, 2000

Versatility of Nuclear Receptors

Transcription factors are proteins that usually function as part of signaling networks in cells. In addition, they can serve as receptors to which hormones or growth factors attach in order to realize their effects on cells. Nuclear receptors comprise a large family of such transcription factors, which can influence whether or not a gene is activated. Thus, nuclear receptors are responsible for many aspects of growth and differentiation.

The biology of nuclear receptors is complex. They are often bound to cellular proteins, called chaperones, which fold the nuclear receptors into certain shapes. They may also be bound directly to DNA. They may interact with binding elements called "ligands," and with different classes of nuclear accessory proteins, called coactivators and corepressors, which foster the ability of the receptor to turn specific genes "on" or "off." Alterations in the nuclear receptor's apparatus can have profound effects on development. These alterations may also play a role in the onset and progression of diseases such as prostate and breast cancer, diabetes, obesity, and osteoporosis. Studies in model organisms, such as the fruit fly and the mouse, have suggested that members of the nuclear receptor superfamily have a long evolutionary history, are well conserved, and constitute a significant portion of the known genomes.

Because of their importance in cellular machinery, nuclear receptors clearly have a major role to play in the development of new drugs to treat disease. For example, certain compounds that mimic the ability of natural dietary lipids to bind to the receptor have been developed. In studies of patients with type 2 diabetes, these drugs enhanced the ability of the nuclear hormone receptor known as PPAR-gamma to function in the fat cell. The patients had increased sensitivity to insulin and alleviation of the insulin resistance found in this form of diabetes.

To enhance the role of nuclear receptors in drug development, researchers seek to devise agents that are specific for only those aspects of the nuclear receptor that need to be targeted in order to treat a particular disease. In that way, they can maximize a drug's effectiveness and reduce unwanted side effects. By focusing on developing drugs that act with high selectivity, researchers have identified a group of agents called "Selective Receptor Modulators," or SRMs. In the case of estrogens, these are

called SERMS, adding the "E" for estrogen. An important example of a SERM is tamoxifen, which binds to the estrogen receptor, blocking its action in estrogen-dependent tumors. The estrogen receptor is found in many tissues and affects reproduction, cognition, and bone density. An ideal SERM would let the estrogen receptor act to maintain bone density, but would not have a stimulatory effect on cell growth in the breast or uterus that could lead to cancer. Certain accessory proteins, such as the "steroid receptor coactivator-2" (SRC-2), have also been found to be over-expressed in some breast tumors. Research suggests that the blocking of SRC-2 may slow or stop the growth of tumors in those patients in whom the coactivator is over-expressed.

The rational design of highly specific new drugs often requires detailed information about the structure of the nuclear receptor—aided by the field of research known as "structural biology." Once a structural biologist has defined the three-dimensional architecture of a molecule—often by crystallization—then other researchers can use computers to design drugs that will "fit" that shape as a glove fits a hand. For example, scientists have created a computer program that identifies candidate drugs based on the structure of one well-characterized complex of a nuclear receptor and its inhibitor. Using this approach, they screened over 150,000 chemical compounds and identified two novel potential inhibitors, which were later shown to be effective. Clearly, the use of computers to perform virtual surveys of hundreds of thousands of drugs could greatly accelerate the evolution of new therapies. By enabling researchers to survey a much larger pool of potential therapeutic agents, computers can identify new or previously unexplored agents, based on the known structure of a nuclear receptor. Advances in drug development based on further insights into the function of nuclear receptors will likely arise from studies of molecules that activate or repress the receptors. For example, researchers have found that the interaction between repressors and their target nuclear receptors is mediated through a relatively small region within the repressors, called a "CoRNR" box, which can now be studied closely to gain a more precise understanding of how repressors function.

The NIDDK's support of research on the nuclear receptor superfamily encompasses many broad, cross-cutting issues. This research has importance for studying not only all the hormones addressed through the field of endocrinology, but also for research on the digestive and urinary tracts, the brain, and other systems of the body. The multi-dimensional nature of research on nuclear receptors is a prime example of the connection between very fundamental studies in the laboratory and patient-oriented research on a wide range of diseases in the clinic. The NIDDK will continue to enhance research on the regulation of nuclear receptors, and other factors of genetic transcription, through research grants and intramural projects. Among the many opportunities to be pursued is the development of a functional atlas of nuclear receptors.

Hu X and Lazar MA. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature* 402(6757):93-96, 1999.

Schapira M, Raaka BM, Samuels HH, and Abagyan R. Rational discovery of novel nuclear hormone receptor antagonists. *Proc Natl Acad Sci USA* 97(3):1008-1013, 2000.

Transporting Needed Proteins

Cellular proteins are transported from one membrane-encased compartment to another by means of membrane-bound containers called vesicles. The first step in the process of transporting new proteins occurs with the bulging outward of the cell membrane. A vesicle then buds from the bulge, is transported, and fuses with the membrane of the target site. Overseeing the final fusion process is a set of proteins called "SNAREs." Scientists have suggested that SNARE proteins are the fundamental means, not only of membrane fusion, but also of delivery of vesicles to a specific target cellular compartment. Recent advances now confirm that SNAREs do indeed serve both functions. The mechanism underlying specificity appears to be a function of molecular structure and binding sites, in that a vesicle from one cellular compartment can only fuse with a specific target compartment. For example, vesicles from the cellular structure known as the endoplasmic reticulum travel only to the Golgi Apparatus.

Interestingly, vesicle fusion with the target site requires that arm-like coils extend from the SNARE protein and bind to both the vesicle and the target at the same time. The complex between the vesicle and the cell surface plasma membrane SNAREs is called a "SNAREpin," which permits a protein product of the

cell to be secreted into the extracellular environment. This fundamental process underlies the secretion of hormones, as well as enzymes and other proteins, so that they can be transported to the sites in the body where they are needed. When the process is abnormal, diseases occur. One example of a disease in which there is a known mis-targeting of the transport process is cystic fibrosis, in which an important protein fails to reach the cell membrane. Another example is the inherited metabolic disorder, oxalosis, in which an enzyme appears to be mis-targeted from one group of intracellular compartments, called peroxisomes, to another location, the mitochondria.

New therapies may derive from this basic research, as scientists find methods to direct or redirect cellular transport mechanisms. In fact, in one fascinating result, investigators have had success in developing a method for oral drug-regulated insulin secretion, using engineered cells in animals. Using a mouse model with elevated blood sugar, they have been able to aggregate proteins in a part of the cellular apparatus called the endoplasmic reticulum, and then use a drug to stimulate the secretion of both insulin and growth hormone. These findings demonstrate that protein secretion can be directly controlled with drugs, and they therefore offer hope of rapid delivery of insulin and other therapeutic compounds.

In other studies of protein transport mechanisms, researchers have found an answer to the question: "How do the cells of the body recycle parts of proteins that are no longer needed or functioning?" Put simply, the answer is "ubiquitin," a protein that is essential for getting rid of cellular clutter. The addition of a ubiquitin "tag" to a protein targets that protein for selective degradation. Ubiquitin is vital in controlling the concentration of key regulatory proteins involved in signal transduction, transcription, and the control of the cell cycle. The "ubiquitin-dependent" protein-degradation pathway is also suspected to play a role in metabolism and the transport of small molecules across membranes. Ubiquitin selectively recognizes a protein via a feature of its structure. The rate of protein degradation may then be regulated by further modification of the protein structure, for example, through addition of a phosphate group to the molecule. In other cases, a modification regulates the activity of one of three enzymes essential to ubiquitin's cell-degrading activity: E1, E2, and E3. The three "E" enzymes are potential drug targets for several diseases, including cervical cancer.

Researchers recently described the regulation of a ubiquitin-dependent pathway through the binding of a chain of two amino acids to a site on the E3 enzyme that is different from the enzyme's active site. The result is to increase the cell's capacity to import peptides. This work established for the first time that the activity of the E3 enzyme can be directly linked to the presence of an environmental signal, such as the peptides from food, in microorganisms. The data also suggest that small compounds may regulate other ubiquitin-dependent pathways.

Because protein degradation by the ubiquitin system controls the intracellular concentrations of many regulatory proteins, it underlies the basic functioning of the cell in health and disease. An example of such regulation is seen in inflammatory bowel disease (see also page 69). The vast majority of microorganisms in the gut do not cause intestinal inflammation. However, the ones that do incite inflammation seem to cause an acute inflammatory colitis by activating a transcription factor that elicits inflammatory proteins called chemokines. Recently, researchers have shown that the interaction of non-disease-causing organisms with the lining of the gut can reduce the synthesis of proinflammatory proteins, by inhibiting the effects of ubiquitin. This newly discovered mechanism appears to explain the unique tolerance of the gastrointestinal lining to pro-inflammatory stimuli. The significance of research on ubiquitin was recently recognized by an Albert Lasker Medical Research Award. In part, this award both illustrates and celebrates the transition of basic research to the clinical bedside, by recognizing that the protein degradation process underlies a broad range of clinical problems.

Fukuda R, McNew JA, Weber T, Parlati F, Engel T, Nickel W, Rothman JE, and Sollner TH. Functional architecture of an intracellular membrane t-SNARE. *Nature* 407(6801):198-202, 2000.

McNew JA, Parlati F, Fukuda R, Johnston RJ, Paz K, Paumet F, Sollner TH, and Rothman JE. Compartmental specificity of cellular membrane fusion encoded in SNARE proteins. *Nature* 407(6801):153-159, 2000.

Neish AS, Gewirtz AT, Zeng H, Young AN, Hobert ME, Karmali V, Rao AS, and Madara JL. Prokaryotic regulation of epithelial responses by inhibition of I-kappa B-alpha ubiquitination. *Science* 289(5484):1560-1563, 2000.

Parlati F, McNew JA, Fukuda R, Miller R, Sollner TH, and Rothman JE. Topological restriction of SNARE-dependent membrane fusion. *Nature* 407(6801):194-198, 2000.

Turner GC, Du F, and Varshavsky A. Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. *Nature* 405(6786):579-583, 2000.

Xavier RJ and Podolsky DK. How to get along–Friendly microbes in a hostile world. *Science* 289(5484):1483-1484, 2000.

Research Tools: Propelling Science Forward

An integral part of scientific inquiry, both informal and formal, is the way in which new information builds upon the existing body of knowledge. As information accumulates, comprehension increases, which, in turn, prompts new questions. A part of this process is the development of research tools to aid the scientific enterprise in the quest for explanations about health and disease. These new tools permit the formulation of questions that were once beyond reach. Through this process, the knowledge base expands to include what was previously unknown and unknowable. Science thus works in a cycle, in which knowledge drives development of technology, and new technology facilitates the gathering of additional insights.

Some advances in research methods represent the adaptation of existing techniques, while others are wholly new. These advances are helping to frame and answer questions about the fundamental mechanisms that underlie normal growth and development, as well as those that promote various disease states. A few are highlighted here.

The Human Genome Project and DNA Microarray Technology: Ten years ago, the determination of the sequence of chemical bases that comprise the human genetic blueprint, deoxyribonucleic acid or DNA, was a time-consuming and tedious laboratory exercise performed at the bench by researchers. After several days of work, one typically could read several hundred bases. Since that time, a major technological advance—the development of large-scale, automated DNA sequencing machines—has enabled the generation of large amounts of DNA sequence information in a reliable and relatively inexpensive manner. The ability to perform many DNA sequencing experiments simultaneously provided researchers with the tools they needed to tackle the ultimate project: sequencing the estimated three billion nucleotides that comprise the chromosomal DNA of humans.

Now, with the Human Genome Project nearing completion, researchers are trying to sort through the resultant trove of knowledge. They must catalog these novel DNA sequences and identify which stretches of DNA are likely to include new genes. Then, they must identify the functions of these genes and begin manipulating this new information in usable ways. These new avenues of research, called bioinformatics and computational biology, will shed light on how genes are turned "on," through biological and/or environmental "switching," to either cause or prevent disease. It is essential to understand this process, called gene "expression," which could then be promoted or inhibited in new ways to treat and prevent the occurrence and progression of disease.

Capitalizing on the explosion of new genetic information for clues about disease will be greatly facilitated by the recent development of new technology. This technology will permit analysis of the data generated by the Human Genome Project—on a much larger scale and more rapidly than was previously possible. One such "enabling" technology is the DNA microarray, which measures the expression of a large number of genes with great efficiency. Microarrays permit scientists to assess the relative expression levels of many genes in a single experiment. The ability for simultaneous assay of hundreds, even thousands, of genes is facilitating the classification and cataloging of novel genes uncovered through the Human Genome Project. While powerful, microarray technology is also expensive. NIDDK is working to make this powerful technology available to a wider range of researchers. A number of new initiatives, including the establishment of Biotechnology Centers, will provide resources to investigators working in research areas within the NIDDK's mission so that they can study the function of genes, efforts termed "gene profiling" or "genomics." Through NIDDK support of such Centers, more investigators will have access to this new technology. For a more complete discussion of microarray technology and its potential applications see the accompanying sidebar on "Enabling Technologies: Microarrays" (page 19).

Clues about Iron Metabolism from a Model Organism: Iron is a mineral that plays an important role in the maintenance of proper health. It has a key role in the synthesis and function of hemoglobin, the oxygen-carrying component of red blood cells. Iron levels must be carefully maintained within a relatively narrow range by an organism.

Too little iron can give rise to anemia, while too much can cause debilitating and possibly fatal iron overload, such as that seen in hemochromatosis. Precisely how the body maintains tight control over iron metabolism is largely unknown. Likely players in this process include not only genes, but also certain proteins that transport iron from the digestive system into the circulation, as well as other proteins that eliminate excess iron from the body.

To gain insights about which genes might play a role in maintaining this delicate balance, scientists studied a model organism, the zebrafish. Zebrafish are attractive models for several reasons, many of which are discussed in the sidebar on "Enabling Technologies: Model Systems" (see page 94). Two aspects of zebrafish biology are particularly relevant for these studies: the relative ease with which their DNA can be manipulated, and the ready observation of their embryonic development, which occurs outside the body of the mother. In order to identify genes that may be involved in iron metabolism, scientists treated a large number of zebrafish embryos with a chemical that damages their DNA, generating random mutations in each embryo. They searched for defects in iron metabolism in two different embryos whose red blood cells lacked red color, indicating the presence of little, if any, hemoglobin. Iron levels in these mutant embryos were four- to nine-fold lower than those seen in normal embryos. Researchers next injected the mutant embryos with iron, after which their blood returned to normal. This finding indicated that the mutant embryos have an impaired capacity to absorb iron, but still retain the ability to use it.

In order to identify the gene or genes altered in these two mutants, scientists analyzed their DNA. Although each mutant embryo carried a unique DNA mutation, in both cases the defect was within the same gene. The protein produced by this gene is thought to be involved in transporting iron from the yolk sac and also to play a role in the development of red blood cells. It is also present in the intestinal tract in adult zebrafish, where it may import iron from the digestive system. In mammals, the protein is present in the intestinal tract, as well as in placenta and liver. Interestingly, all these tissues are known to be prominent sites of iron transport. Scientists named this gene *ferroportin 1*, combining the words "ferro," meaning iron, and "portin," meaning transport. Future experiments will address the question of whether perturbations in *ferroportin 1* in humans may be responsible for diseases of iron deficiency or overload.

As part of a trans-NIH effort, the NIDDK is supporting studies of the zebrafish genome. Just as the information from the Human Genome Project is revolutionizing the study of human disease, the knowledge of the zebrafish genome will facilitate understanding of this powerful model organism as research advances are translated into new therapies. Research initiatives will seek to identify which genes are expressed in the zebrafish and to create an atlas of gene expression indicating which genes are expressed in which tissues and at what times during development. The NIDDK will also support efforts to enhance mutagenesis studies of zebrafish such as the one described previously. Such studies, in which genes are mutated experimentally in animal models, are an important way to gain a better understanding of how a gene can cause disease and what interventions could possibly arrest that process. Because the zebrafish is a vertebrate animal, unlike other genetic models such as worms and fruit flies, information gleaned from its study may be more directly applicable to human conditions. Facilitating advances in this area of study may therefore lead to significantly enhanced understanding of human disease.

A Useful Technique To Study Signaling Within a Cell: Genes form a DNA blueprint from which proteins are constructed. Genes that are turned "on" and give rise to mature proteins are said to be "expressed." Genes do not function in a vacuum, however. Rather, the maintenance of health requires highly precise coordination in the expression of hundreds if not thousands of genes. Defects in relatively few genes, or even in a single one, can cause problems at the level of cells, tissues, and organs and lead to a host of diseases and their complications. As noted previously, zebrafish studies have shown how a mutation in a single gene that encodes a protein present in a limited range of tissues can nevertheless have consequences that negatively impact the well-being of the entire organism. It is therefore important for scientists to assess the role played by any one gene in the context of a complex cellular milieu. They often seek to inhibit expression of a single gene with high selectivity, and then to determine the impact of the gene's functional absence on cellular metabolism and function. However, many drugs used to inhibit gene expression are relatively nonspecific, and it is often difficult to fully inhibit expression

of only the gene under study, while leaving others unaffected. Conversely, other agents that act in a relatively specific manner to inhibit gene expression may do so incompletely, which makes it difficult to achieve complete ablation of gene expression.

Helping to resolve some of these problems is a technique called RNA-mediated genetic interference (RNAi). It involves the use of synthetic double-stranded RNA to inhibit gene expression, through a mechanism not yet understood. However, researchers do know that RNAi is highly specific, effectively inhibits gene expression, and seems to have minimal side effects on target cells. This technique has been widely used by researchers studying the roundworm *C. elegans*. In these experiments, the worms are either directly injected with double-stranded RNA molecules or are maintained in a solution containing double-stranded RNA. Researchers have also used RNAi to inhibit gene expression in the fruit fly *Drosophila* by injecting double-stranded RNA into developing embryos. While both approaches have yielded important information, they do suffer from limitations.

Genes do not function in a vacuum. Rather, the maintenance of health requires highly precise coordination in the expression of hundreds if not thousands of genes.

Scientists have recently described the adaptation of RNAi to inhibit gene expression in *Drosophila* cells grown in culture. This advance offers many experimental advantages, both scientific, in terms of its flexibility, and practical, because the double-stranded RNA does not need to be injected and can simply be added to the growth medium. Recently, NIDDK-supported researchers demonstrated the validity of this approach by using RNAi to selectively reduce the expression levels of a number of proteins in *Drosophila* cultures, including several that act in the insulin-signaling pathway. RNAi decreased expression of targeted genes to only one-to-five percent of normal levels. This decrease in protein levels was sufficient to inhibit the activation of downstream signaling molecules in the insulin-signaling cascade, thus demonstrating that this approach can be used to remove the function of a specific gene's protein.

The adaptation of RNAi to use in culture offers several advantages over traditional methods for inhibiting gene expression. It is technically straightforward and highly reproducible. The use of RNAi in fruit flies is of particular significance, because it will aid in the characterization of a number of biochemical pathways that are similar to those found in higher organisms. For example, the fruit fly's insulin-signaling pathway is very similar to that of humans. Furthermore, a large amount of genetic information already exists regarding the fruit fly's genome, both as a result of longstanding efforts of the research community, as well as fly "genome projects," which may also facilitate the cataloging of many genes uncovered as part of the Human Genome Project. These experiments may also serve as a stepping stone for adaptation of this technique to mammalian systems. It offers a much faster alternative to traditional methods of generating gene deletions, called gene "knockouts" in mammalian cells, which can require weeks or months of preparation.

New Approaches To Generating Tissue-Specific Knockouts: A powerful tool is the ability to inhibit gene expression with higher specificity in culture; however, this approach is not without limitations. Some questions cannot be answered by experiments on cells growing in culture. For example, studies of embryonic development, or of how organ systems communicate and coordinate with one another, must be performed in the context of a whole organism. Scientists have therefore devoted much time and energy to studying animals, whether normal or genetically modified, to understand the role played by specific genes. However, producing "knockout" animals, in which one or more genes are specifically deleted, is labor-intensive and time-consuming. Moreover, there is no guarantee that such efforts will produce a definitive answer. Many genes are members of multi-gene families that perform related functions. Therefore, deletion of a single gene may have little impact on the overall health of the animal if the remaining genes can effectively compensate for the lost one. Alternatively, deletion of genes that play critical roles in the formation and growth of the embryo can cause severe developmental defects or even prenatal death, precluding study of the gene's function in subsequent biological processes. One way to get around this problem is to generate "conditional" gene knockouts, in which the gene of interest is present during early development, but is subsequently deleted. With this approach, it is possible to produce mice which lose a target gene at a specific point in time, in specific tissues, or both.

NIDDK-supported researchers have recently described two methods of generating time-specific or tissue-specific gene knockouts in mice. These scientists targeted the cells that line the small intestine and colon, called epithelial cells. With both methods, the gene of interest remains intact and is expressed normally until the appropriate signal is received. In the first approach, this signal is triggered by an endogenous event, such as when the cells reach a specific stage of development. Using this technique, researchers were able to delete a target gene in developing intestinal cells at day 14 of embryonic development. A few days later, the gene was deleted from cells in the kidney, pelvis, uterus, and bladder, as they reached the same developmental stage. This approach relies on biologically-derived signals generated by the target cells themselves. In a second, similar experiment, researchers used a slightly modified approach to delete a gene from these same tissues in response to an exogenous event, over which the researchers exerted control; namely, the administration of the antibiotic doxycycline. This approach allows the gene to be deleted at any time through the simple addition of a drug to the drinking water of experimental animals. It therefore allows researchers to choose a specific time at which the gene of interest would be deleted independent of the developmental stage of the animals.

The specificity this technology brings to research studies is an important advance for scientists who want to study gene function through genetic knockouts, but cannot do so because of problems such as lethality. Importantly, this strategy, with minor modifications, is potentially applicable to any gene and any tissue. The NIDDK supports a number of broad-based initiatives that are aimed at the study of mouse models.

Genetic factors are thought to underlie many diseases within the NIDDK mission, including diabetes and its complications, and obesity. Mutant mouse models have been generated by a large number of investigators to study these processes. However, many of these mouse models exhibit very small defects under normal conditions and only develop disorders under specific circumstances. The NIDDK will therefore establish several Mouse Metabolic Phenotyping Centers in order to make available to researchers the facilities they need to perform detailed analysis of mouse models.

The NIDDK is also assembling a cross-disciplinary consortium to develop innovative mouse models of diabetes complications that closely mimic human disease. Current mouse models of diabetes typically focus on initial development of the disease. In contrast, the pathways leading to subsequent organ damage have received less attention. The initiative on Mouse Models of Diabetic Complications has two goals: (1) to generate animal models of disease development, prevention, and treatment; and (2) to test the role of candidate genes identified by genetic studies of diabetes in humans. Through this effort, the NIDDK hopes to spur research in this critical area of diabetes. The Institute is also supporting a number of other initiatives that exploit the power of mice as model systems.

New Insights into Gene Expression: The process by which specialized cells emerge from unspecialized precursors is known as differentiation. The many tissues and organs that perform diverse functions within the complete organism arise from a small, initially amorphous group of identical cells. Cell differentiation also occurs in the adult, where certain cell populations give rise to specialized daughter cells. These self-renewing "stem cells" have been the focus of much research, because it is thought that a better understanding of the molecular signals that control differentiation may hold important clues regarding a number of diseases. Scientists are therefore very interested in studying the dynamic changes in gene expression that accompany differentiation.

In order to study gene expression closely, researchers ideally would like to identify a single stem cell and its lineage—all the daughter cells arising from it. However, stem cells and their progeny usually exist within populations of cells. Thus, it has previously been difficult to establish definitively which differentiated cells arise from which progenitors. Furthermore, the number of cells at any specific point in the differentiation pathway is usually quite small, and these cells are often intermingled with cells at other stages of differentiation or with different cell types entirely. Questions about cell lineage are also relevant beyond the topic of stem cells. Many diseases are focal in origin; they begin not as diffuse disorders, but as malfunctions within a single cell or small group of cells. Understanding cell lineage may therefore provide insights into the origin and evolution of such diseases.

NIDDK-supported scientists have recently described a model system that addresses both of these problems. They have devised a method of generating mice bearing a

molecular identifier—a so-called "marker"—within a subset of stem cells in their intestinal tracts. This marker is present within the stem cells that give rise to villi—tiny hairlike projections that consist of constantly renewing cells that line the inside of the small intestine. As a result, the marker identifies all daughter cells that arise from a particular progenitor. By isolating and treating this tissue with the appropriate chemicals, scientists can cause "marked" cells to turn blue, making them easily identifiable. Importantly, this marker gene is expressed in a mosaic fashion, that is, some but not all of the stem cells have it. This mosaic expression pattern allows scientists to observe a mature villus and determine which cells are derived from a particular marked stem cell, because these daughter cells will all stain blue, while cells from unmarked stem cells will not. With this technique, scientists can clearly identify cells of a unique lineage within a villus, study the fates of cells of different lineages, and study interactions with cells of other lineages.

In addition to identifying cells of a particular lineage in the gut, these scientists also have isolated small numbers of cells from a single lineage. They modified and applied a previously described technique known as laser capture microdissection, LCM. To perform LCM, scientists embed a tissue sample in a semi-solid matrix and then use a machine to slice it into very thin layers. These tissue sections are placed on a microscope slide and a thin transparent film is placed over them. When the researcher directs a low-intensity laser beam at specific cells, they heat slightly and adhere to the overlying film, detaching from the rest of the sample. Cells so selected are removed, and may be used for a number of studies. Unfortunately, the chemical manipulations necessary to identify cells carrying the marker also destroy the genetic material within the cells. This limitation had rendered this technique unsuitable for studying cell lineages until the recently introduced modification by NIDDK-supported scientists.

The modification provides a way to identify marked cells without having to stain them, thereby preserving their genetic material for subsequent analysis. To do this, researchers made tandem slides using adjacent tissue sections that differ very little and contain many of the same cells. After staining one of the slides, they used it to generate a computerized "map" of the section, including cells that turned blue. They then used this map to guide their LCM of the adjacent, unstained section to precisely isolate only those cells from a given lineage. Because these cells had not been stained, their genetic material was intact. Using this technique, the researchers were able to specifically separate marked from unmarked cells and to perform genetic analysis on the isolated cells. The development of this model system and modified LCM technique may facilitate further research into cell lineage within the cells of the gut. It may also lead to significant advances in numerous digestive diseases, especially in those thought to arise from defects within a single cell.

Measuring Protein Size: Errors in protein composition are responsible for many disorders and diseases. Scientists study DNA because it represents the relatively unchanging template from which proteins are derived; the ultimate product of gene expression is a mature protein. Each protein is made up of a unique combination of amino acids, and each has a unique size. Proteins themselves are a critical target for investigation because their proper function requires more than just faithful translation of the instructions in DNA. After synthesis, proteins may undergo a series of modifications. As a result, detailed knowledge of a protein's size is an important tool in identifying it and verifying its proper components. One group of proteins that has proven resistant to most traditional approaches to size measurement is "transmembrane" or "membrane-bound" proteins, which comprise about ten percent of all proteins within a cell. These proteins are partially embedded in the layer of fat that makes up cellular membranes. They can be receptors that detect and transmit important cell signals; act as channels that allow nutrients, ions, and waste products to cross in and out of the cell; and perform many other important cellular functions.

NIDDK-supported investigators have developed a widely applicable method that permits rapid determination of the molecular mass of a full-length protein with high precision. Investigators have used this technique successfully to perform analysis of four different membrane proteins from three bacterial organisms. This approach can be used to identify novel proteins; to detect small errors in the amino acid sequence of known proteins by precisely measuring their size; and to characterize the nature and extent of modifications to proteins that occur after this assembly. Importantly, this new technique can be used to identify proteins within a crude tissue preparation. Consequently, it eliminates the need for extensive purification, which is often technically diffi-

cult and time-consuming. This new technique can provide information necessary for determining a protein's structure and function—something that will be especially useful when classical methods for measuring a protein's size have failed.

The completion of the Human Genome Project, as well as the continued sequencing of a number of other genomes, will lead to the identification of many genes and their function. As understanding of various genomes has grown, the study of organisms has evolved from genetics, the study of genes, to "functional genomics," the study of how all the genes of an organism work together and coordinate to perform all the functions necessary for life. A similar evolution is occurring in the study of proteins. Proteomics, the study of how proteins work together, is an emerging field. The development of a new technique for measuring the size of a protein, even when it is contained within a mixture of many cellular proteins, will allow rapid analysis of proteins produced by novel genes. The NIDDK is supporting advances in the proteomics field through an expansion of the Institute's intramural structural biology efforts and through enhanced support for extramural research projects.

Clemens JC, Worby CA, Simonson-Leff N, Muda M, Maehama T, Hemmings BA, and Dixon JE. Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc Natl Acad Sci USA* 97(12):6499-6503, 2000.

le Coutre J, Whitelegge JP, Gross A, Turk E, Wright EM, Kaback HR, and Faull KF. Proteomics on full-length membrane proteins using mass spectrometry. *Biochemistry* 39(15):4237-4242, 2000.

Donovan A, Brownlie A, Zhou Y, Shepard J, Pratt SJ, Moynihan J, Paw BH, Drejer A, Barut B, Zapata A, Law TC, Brugnara C, Lux SE, Pinkus GS, Pinkus JL, Kingsley PD, Palis J, Fleming MD, Andrews NC, and Zon LI. Positional cloning of zebrafish *ferroportin1* identifies a conserved vertebrate iron exporter. *Nature* 403(6771):776-781, 2000.

Saam JR and Gordon JI. Inducible gene knockouts in the small intestinal and colonic epithelium. *J Biol Chem* 274(53):38071-38082, 1999.

Wong MH, Saam JR, Stappenbeck TS, Rexer CH, and Gordon JI. Genetic mosaic analysis based on Cre recombinase and navigated laser capture microdissection. *Proc Natl Acad Sci USA* 97(23):12601-12606, 2000.

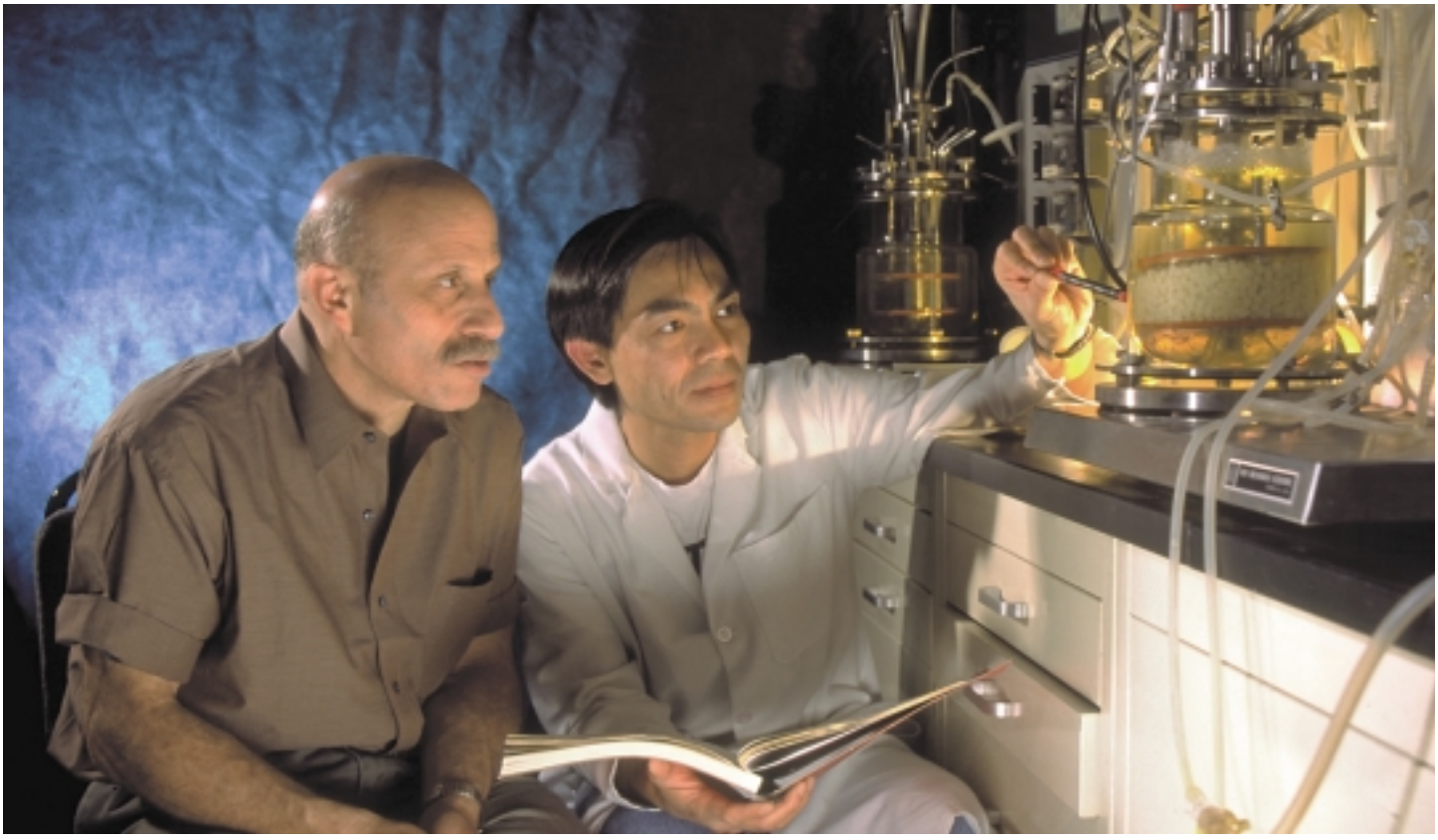
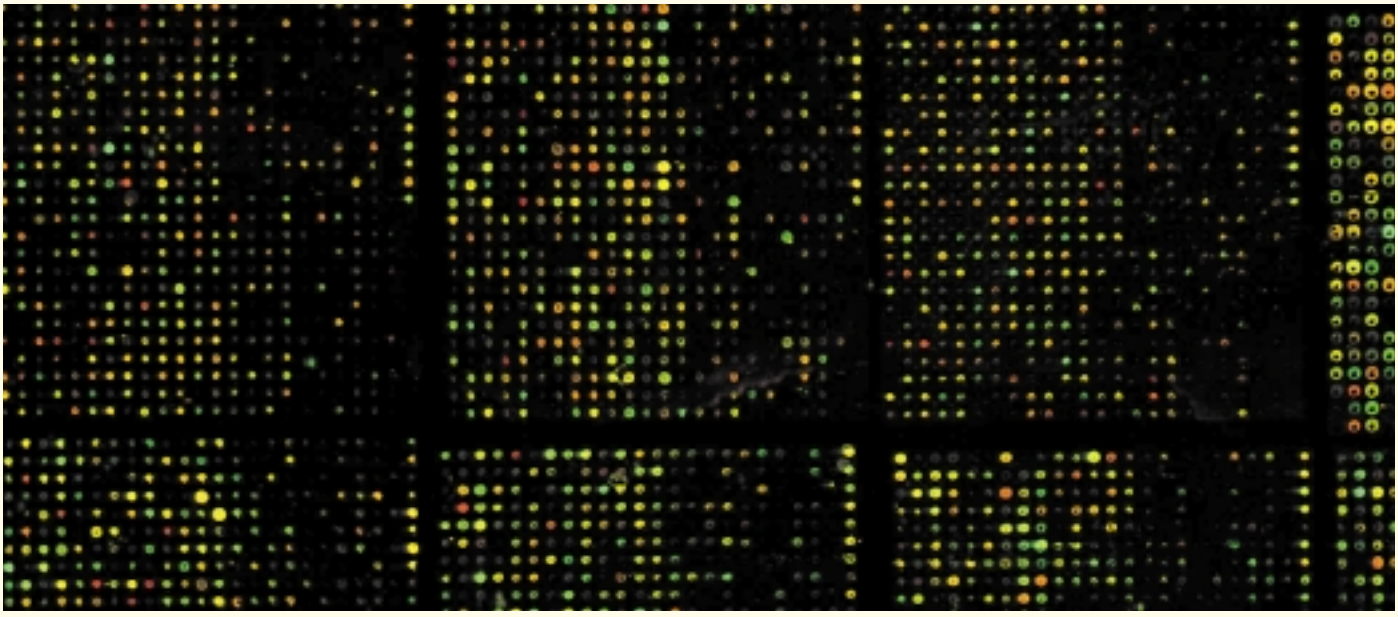


Photo: Mr. Richard Nowitz.

Microarrays

Biomedical research evolves and advances not only through the compilation of knowledge, but also the development of new technologies. Using traditional methods to assay gene expression, researchers were able to survey a relatively small number of genes at a time. The emergence of new tools enables researchers to address previously intractable problems and to uncover novel potential targets for therapies. Microarrays allow scientists to analyze expression of many genes in a single experiment quickly and efficiently. They represent a major methodological advance and illustrate how the advent of new technologies provides powerful tools for researchers. Scientists are using microarray technology to try to understand fundamental aspects of growth and development as well as to explore the underlying genetic causes of many human diseases, including diabetes, bowel disease, and kidney disease.

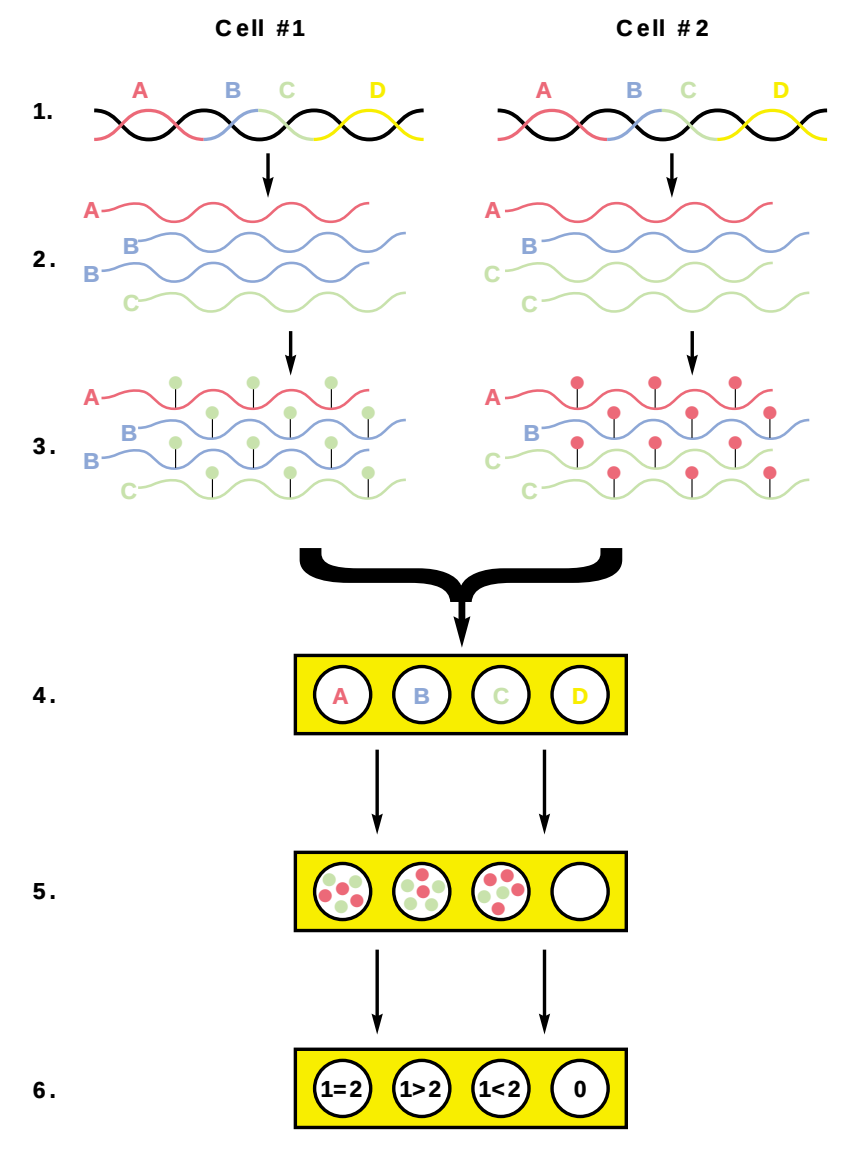
With only a few exceptions, every cell of the body contains a full set of chromosomes and identical genes. Only a fraction of these genes is turned on, however, and it is the subset that is "expressed" that confers unique properties to each cell type. "Gene expression" is the term used to describe the transcription of the information contained within the DNA, the repository of genetic information, into messenger RNA (mRNA), molecules that are in turn translated into the proteins that perform most of the critical functions of cells. Scientists study the kinds and amounts of mRNA produced by a cell to learn which genes are expressed, generating knowledge which, in turn, provides insights into how the cell acts. Gene expression is a highly complex and tightly regulated process that allows a cell to respond dynamically both to environmental stimuli and to its own changing needs. The gene expression process is not only an "on/off" switch to control which genes are expressed in a cell, but it is also a "volume control" that increases or decreases the level of expression of particular genes as necessary. The proper and harmonious expression of a large number of genes is a critical component of normal growth and development and the maintenance of proper health. Disruptions or perturbations in gene expression are responsible for many diseases.



Microarrays, like the one shown above, use fluorescent markers to show which genes in a cell are actively expressed and making proteins. The method uses precisely applied tiny droplets of functional DNA on a glass slide (shown above), silicon chip, or membrane. Researchers then attach fluorescent labels to genetic material from the cell or cells that they are studying. The labeled probes are allowed to bind to complementary DNA strands on the slides. The slides are put into a scanning microscope that, with the help of a computer, measures the brightness of each fluorescent dot; brightness reveals how much of a specific DNA fragment is present, an indicator of how active it is. Photo: National Human Genome Research Institute, National Institutes of Health.

Two recent complementary advances, one in knowledge and one in technology, are greatly facilitating the study of gene expression and the discovery of the roles played by specific genes in the development of disease. As a result of the Human Genome Project, there has been an explosion in the amount of information available about the DNA sequence of the human genome. Consequently, researchers have identified within these previously unknown sequences a large number of novel genes. The challenge currently facing scientists is to find a way to organize and catalog this vast amount of information into a usable form. Only after the functions of the new genes are discovered will the full impact of the human genome project be realized.

The second advance may facilitate the identification and classification of this DNA sequence information and the assignment of functions to these new genes: the emergence of DNA microarray technology. A microarray is a tool for analyzing gene expression that consists of a small membrane or glass slide containing samples of many genes arranged in a regular pattern. A microarray works by exploiting the ability of a given mRNA molecule to bind specifically to the DNA template from which it originated. By using an array containing many DNA samples, scientists can determine—in a single experiment—the expression levels of hundreds or thousands of genes within a cell by measuring the amount of genetic material bound to each site on the array. With the aid of a computer, the amount of labeled genetic material bound to the spots on the microarray is precisely measured, generating a profile of gene expression in the cell. Microarrays are a significant advance both because they may contain a very large number of genes and because of their small size. Microarrays are therefore useful when one wants to survey a large number of genes quickly or when the sample to be studied is small. Microarrays may be used to assay gene expression within a single tissue or cell



Consider two cells: Both contain an identical set of genes, A, B, C, and D (1). Scientists are interested in determining the profile of expression of these four genes in two cell types. To do this, they isolate mRNA from each cell type (2) and use these molecules as templates to produce molecules with a molecular "tag" attached (3). In this illustration, two different tags are used, so that the samples can be differentiated in subsequent steps. These two labeled samples are mixed and incubated with a microarray containing genes A, B, C, and D (4). The labeled molecules bind to the sites on the array corresponding to the genes expressed in each cell (5). The amount of bound labeled molecules is determined using a computer. In this example, the computer concludes that both cells express gene A at the same level, that cell 1 expresses more of gene B, that cell 2 expresses more of gene C, and that neither cell expresses gene D (6). In a real microarray experiment, the expression of thousands of genes would be compared.

type as a function of some treatment or developmental change, or to compare gene expression in two different cell types or tissue samples. NIDDK-funded scientists are using microarrays to investigate a wide range of diseases. The application of this exciting new technology has resulted in the following recent reports.

Inflammatory Bowel Disease (IBD): Nearly one million Americans are afflicted with the inflammatory bowel diseases of Crohn's disease and ulcerative colitis. IBD is a debilitating autoimmune disease causing fever, abdominal pain, diarrhea, weight loss, and fatigue. It is believed to arise from an inappropriate reaction by the body's immune system to the normally harmless bacteria in the gut, resulting in chronic inflammation and tissue damage. To identify genes involved in mediating the inflammatory response, scientists studied gene expression in cultured human colorectal cells infected with the bacterium *Salmonella*. These investigators used a microarray containing over 4,000 genes and found that, surprisingly, only a small number of genes changed expression following infection. This group included both known and previously undescribed genes. Thus, this application of microarray technology suggests that only a relatively small group of genes may be critical mediators of the immune response in cells of the gut, and may play a role in the underlying disease process in patients with IBD.

Type 1 Diabetes: Type 1 diabetes results from the destruction of the insulin-producing beta cells of the pancreas by the immune system. To shed light on the unknown genetic causes of type 1 diabetes, scientists study "discordant" twins—identical twins in which one develops type 1 diabetes while the other does not. By doing so, they hope to identify subtle genetic changes that may play a role in disease development—such as differences in the cells of the immune system. Research suggests that type 1 diabetes may arise, in part, from an imbalance between two subtypes of immune cells known as T cells. T cells respond to potentially harmful agents by rearranging specific regions of chromosomal DNA to fight infection. While identical twins are born with the same genetic information, their DNA will not remain 100 percent identical over time because they may be exposed to different environmental stimuli or they may respond to the same stimulus in slightly different ways. In many discordant twins, T cells in the normal twin produce a protein called interleukin-4 (IL-4), while T cells in the diabetic twin produce less IL-4. Researchers used a microarray containing nearly 7,000 genes to compare gene expression in T cells from a pair of discordant identical twins. In addition to differences in expression of IL-4, scientists found that regulation of many other genes also differed between the twins. Taken as a whole, the pattern of gene expression suggests that T cells from the diabetic twin are less mature than those of the normal twin, and that this difference may partly explain the tendency of these cells to attack pancreatic beta cells. Future studies will attempt to identify the specific roles that these differentially expressed genes may play in the development of type 1 diabetes.

Type 2 Diabetes: People with type 2 diabetes often continue to produce insulin, but the cells of their bodies no longer respond to it properly. Ultimately, the beta cells of the pancreas may fail, causing insulin to drop to dangerously low levels. "Glucose toxicity" is the term used to describe the damage to beta cells caused by prolonged exposure to elevated blood glucose levels. To understand how glucose toxicity negatively impacts beta cells, scientists used a microarray to compare gene expression in cultured beta cells grown under low or high glucose conditions. Out of over 6,000 genes surveyed, approximately 80 changed expression levels under different glucose concentrations. Perhaps not surprisingly, the two largest groups of genes with altered expression were those involved in secretion and in cellular metabolism, processes known to be abnormal in type 2 diabetes. Importantly, genes involved in protein metabolism were expressed under low glucose conditions. This finding suggests that these molecules provide an important source of energy for the beta cell when glucose is scarce. Through the use of microarrays, scientists are gaining a fuller picture of the genetic abnormalities associated with type 2 diabetes.

Hematopoietic Stem Cells: "Stem cells" are progenitor cells from which the highly specialized cells and tissues of the body develop. In adults, an important stem cell found primarily in the bone marrow is the hematopoietic stem cell. These cells give rise to a variety of blood cell types. Researchers are trying to uncover the molecular mechanisms by which stem cells remain in an undifferentiated state or develop stepwise into mature cells. They also want to understand why stem cells present during the earliest stages of development in animals seem to be more "plastic," in that they can differentiate into a wider variety of cell types, than stem cells in adults. Researchers have therefore used microarrays to compare gene expression in hematopoietic stem cells isolated from fetal and adult mice to gain a better comprehension of the molecular basis of these differences. One study that examined about 18,000 genes revealed numerous genes whose expression is significantly different in the two cell populations. Researchers identified a large number of genes that could potentially play a role in controlling stem cell growth and differentiation. They also identified a large number of previously undescribed genes whose roles are unknown. Importantly, cohorts of genes were identified that correlated with specific stages of stem cell differentiation, allowing researchers to generate a genetic "profile" of genes expressed at a given stage. Thus, microarray technology enabled the creation of a database of gene expression that is an important step in the characterization of the molecular basis of stem cell growth and differentiation. Insights gained from this database may be important to future understanding of blood disorders, as well as for a wide range of other diseases for which hematopoietic stem cells offer therapeutic promise.

Kidney Diseases: The kidney plays a vitally important role in ensuring the health of an organism by filtering blood and removing waste products that could otherwise be poisonous. The kidney consists of a complex series of tubules and branches comprised of several types of highly specialized cells, each of which plays a distinct role in the removal of waste and toxins. Researchers are increasing knowledge about the precise roles played by these various cell types, and how they interact with one another. They have applied DNA microarray technology to generate a profile of all genes expressed in human kidney. Using an array containing 18,000 human genes, they identified over 7,500 genes expressed in normal human kidney cortex. In a related study, scientists used a microarray containing nearly 600 genes to study kidney cells grown in culture that, under appropriate conditions, form the complex branched tubules known as the ureteric buds. These investigators found that cells induced to undergo branching expressed different genes than those that were not, and they also defined subsets of genes expressed at distinct times during the branching process. These studies represent two early indices of gene expression in kidney, both under normal conditions and during branching, and will be important tools in understanding the growth and development of this complex organ.

Because a microarray can be used to examine the expression of hundreds or thousands of genes at once, it promises to revolutionize the way scientists examine gene expression. This technology is still in its infancy; therefore, many initial studies using microarrays have represented simple surveys of gene expression profiles in a variety of cell types. Nevertheless, these studies

represent an important and necessary first step in the understanding and cataloging of the human genome. As more information accumulates, scientists will be able to use microarrays to ask increasingly complex questions and perform more intricate experiments. With each new advance, researchers will be able to infer probable functions of new genes based on similarities in expression patterns with those of known genes. Ultimately, these studies promise to expand the size of existing gene families; reveal new patterns of coordinated gene expression across gene families; and uncover entirely new categories of genes. Furthermore, because the product of any one gene usually interacts with those of many others, perception of how these genes coordinate will become clearer through such analyses, and precise knowledge of these inter-relationships will emerge. The use of microarrays may also speed the identification of genes involved in the development of various diseases by enabling scientists to examine a much larger number of genes. This technology will also aid the examination of the integration of gene expression and function at the cellular level, revealing how multiple gene products work together to produce physical and chemical responses to cellular needs.

This new microarray technology and the information it generates present both great opportunities and great challenges for the biomedical research community today. One challenge facing researchers is the high cost of using this new technology to investigate problems of interest. The NIDDK has recognized this problem and has devised a strategy to address it. A new initiative will make funds available to grantees to allow them to purchase the specialized equipment necessary to perform microarray analysis. In addition, the Institute is promoting the development of "core facilities" whose resources would be shared by several NIDDK-supported investigators in an attempt to place this new technology into the hands of more researchers. On a more long-term basis, a trans-NIDDK strategic planning group on "Genetics, Genomics, and Bioinformatics" will provide advice and guidance regarding avenues of future investigations. Through these and other efforts, the NIDDK is taking a proactive role to ensure its research community can fully exploit the host of powerful enabling technologies that are moving the biomedical research enterprise forward in the 21st century.

Eckmann L, Smith JR, Housley MP, Dwinell MB and Kagnoff MF. Analysis by high density cDNA arrays of altered gene expression in human intestinal epithelial cells in response to infection with the invasive enteric bacteria *Salmonella*. *J Biol Chem* 275(19):14084-14094, 2000.

Pavlova A, Stuart RO, Pohl M, and Nigam SK. Evolution of gene expression patterns in a model of branching morphogenesis. *Am J Physiol* 277(4 Pt 2):F650-663, 1999.

Phillips RL, Ernst, RE, Brunk, B, Ivanova N, Mahan MA, Deanehan JK, Moore KA, Overton GC, and Lemischka IR. The genetic program of hematopoietic stem cells. *Science* 288(5471):1635-1640, 2000.

Webb GC, Akbar MS, Zhao C, and Steiner DF. Expression profiling of pancreatic beta cells: Glucose regulation of secretory and metabolic pathway genes. *Proc Natl Acad Sci USA* 97(l1):5773-5778, 2000.

Wilson SB, Kent S, Horton HF, Hill AA, Bollyky PL, Hafler DA, Strominger JL, and Byrne MC. Multiple differences in gene expression in regulatory V-alpha-24J-alpha-Q T cells from identical twins discordant for type 1 diabetes. *Proc Natl Acad Sci USA* 97(13):7411-7416, 2000.

Yano N, Endoh M, Fadden K, Yamashita H, Kane A, Sakai H, and Rifai A. Comprehensive gene expression profile of the adult human renal cortex: Analysis by cDNA array hybridization. *Kidney Int* 57(4):1452-1459, 2000.